Hiding in Plain Sight:
A Measurement and Analysis of Kids' Exposure to Malicious URLs on YouTube

Sultan Alshamrani[1], Ahmed Abusnaina[1], David Mohaisen[1]

[1] University of Central Florida

This work was supported by

The Third ACM/IEEE Workshop on Hot Topics on Web of Things, IEEE HotWoT 2020
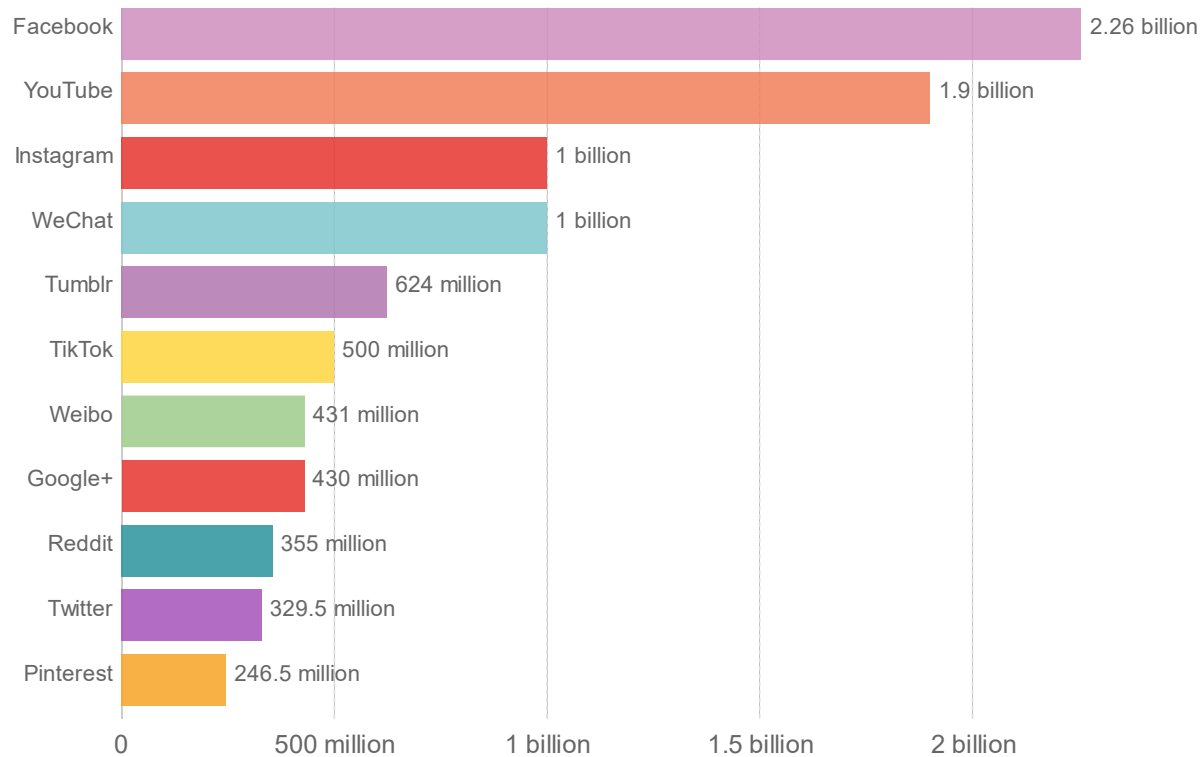
UCF

Outline

- Introduction and Motivation

- Contribution

- Data Collection

- URL Extraction

- Results and Findings

- Conclusion Remarks

UCF

Introduction: Social Media

Number of people using social media platforms, 2018

Estimates correspond to monthly active users (MAUs). Facebook, for example, measures MAUs as users that have logged in during the past 30 days. See source for more details.

Our World in Data

Platform	Users
Facebook	2.26 billion
YouTube	1.9 billion
Instagram	1 billion
WeChat	1 billion
Tumblr	624 million
TikTok	500 million
Weibo	431 million
Google+	430 million
Reddit	355 million
Twitter	329.5 million
Pinterest	246.5 million

0   500 million   1 billion   1.5 billion   2 billion

Source: Statista and TNW (2019)

UCF

Kids on Social Media

Nearly one in five teenagers is spending 4 hours or more online every day

How many minutes, if any, would you say that you spend online (e.g., browsing the internet, watching videos, playing games, chatting with friends, etc.) in a typical day? (%)

■ Children (ages 8 - 12)   ■ Teens (ages 13 - 17)

	None	1 minute to less than 1 hour	1 hour to less than 2 hours	2 hours to less than 3 hours	3 hours to less than 4 hours	4 hours or more
Children	4	31	26	19	7	14
Teens	0	38	16	16	12	17

YouGov

September 6 - 12, 2019

Introduction: Users' Interaction

- YouTube provide users with interactive options such as
    - likes, dislikes as well as commenting.

UCF

Introduction: Users' Interaction

- Commenting has allowed some users to post
  - Malicious URLs.
  - URLs to inappropriate website.

Motivation

- Such inappropriate URLs can be targeted towards users irrespective of age.

- Kids may intentionally or accidentally access the content of the URLs.

Contribution

- We collected around 4 million comments posted on children's YouTube videos.

- An in-depth analysis of kids' exposure to malicious URLs.

Contribution

- From 8,677 URLs, studied the URLs associated topics and audience interaction with inappropriate websites, such as illegal content and adult websites.

- We report on several malicious URLs detected by VirusTotal.

The Selected Kids' Videos

- Top-200 children's shows based on Ranker.

- The list of shows was originally made by Ranker TV and received more than 1.2M votes and has 380 kids' shows In which we selected the top 200 shows.

Ranker is a crowdsourced platform that relies on millions of users to rank a variety of media contents such as shows and films.

Age Assignment

- We mainly used Common Sense Media as the main source for defining the age group of the selected children's shows.

- For the shows that are in not Common Sense Media, we used Parents Guide in IMDB to get the appropriate age.

Common Sense Media is a non-profit organization that provides education and advocacy to families on providing safe media for children

Collection Approach

- 200 shows to YouTube Video API and retrieved the top-50 videos.

- Using video's ID to obtain video statistics, such as views, likes, dislikes, etc.

- Utilizing Comments API to collect all comments from the videos.

UCF

Data Statistics and Measurements
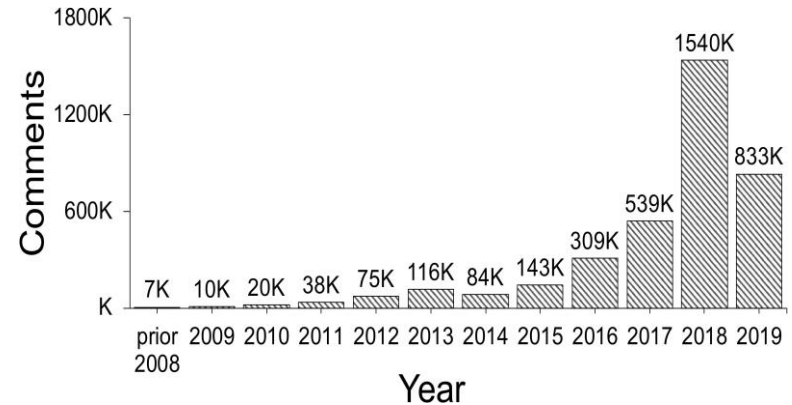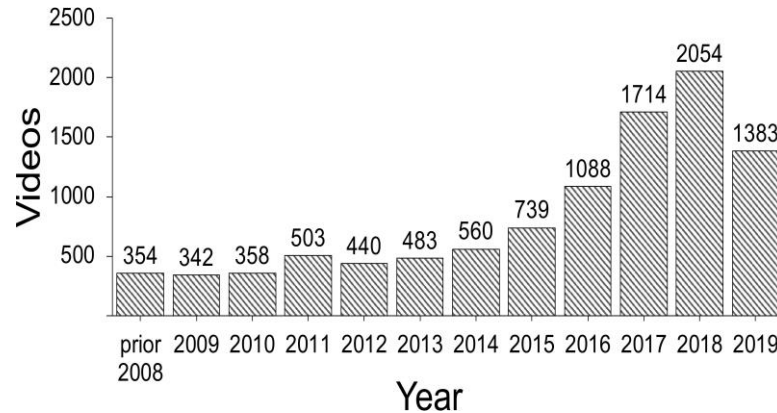
- Rapid increase in children's videos over the past few years thus increase in the number of comments.

- The comments were posted by more than 2.5 million users on about 10,000 videos from ≈3,000 different channels.

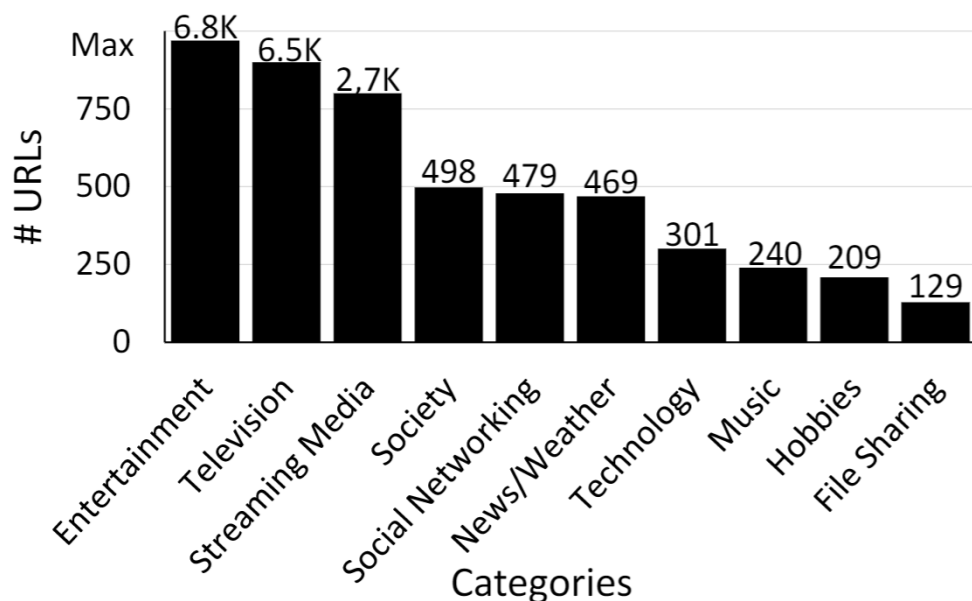- The average viewers count is roughly 2.4 million views and the average comments count is 8,068 comments per video.

URL Extraction

- We used a regular expression to extract URLs within the comments.

- In the collected dataset, we extracted 8,677 URLs.
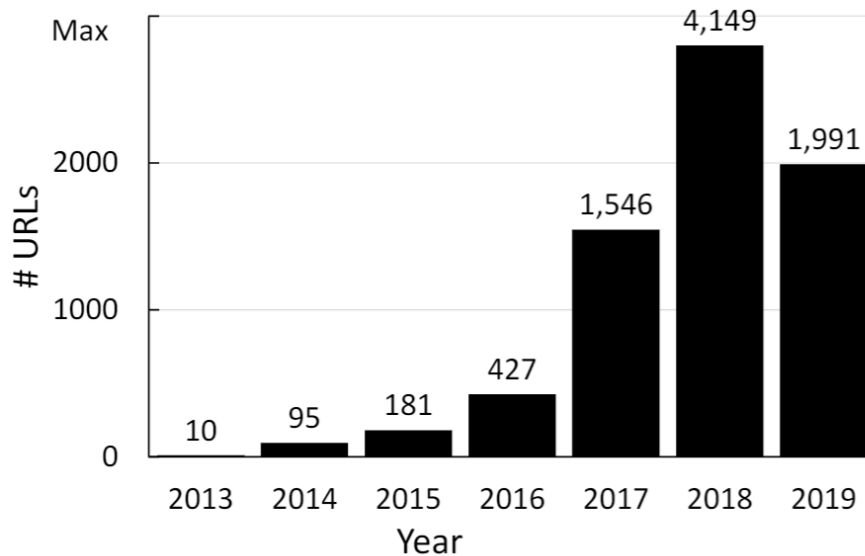
URL Topic Categorization

- Using Webshrinker, we extracted 107 different categories associated with the URLs.

A machine learning-powered domain data, and threat classifier, to obtain the Interactive Advertising Bureau (IAB) categorization of the domains of the URLs

Malicious URL Extraction

- Checked URL is valid or not then forward the URL to VirusTotal API to check whether it is benign or malicious.

VirusTotal
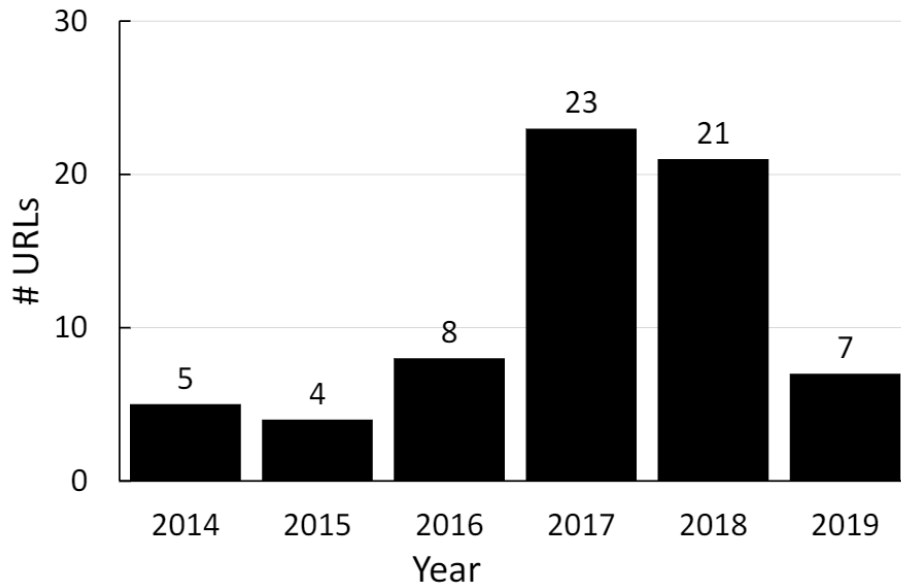
A website aggregates many antivirus products and online scan engine to detect for malicious file and URL analyzer.

UCF

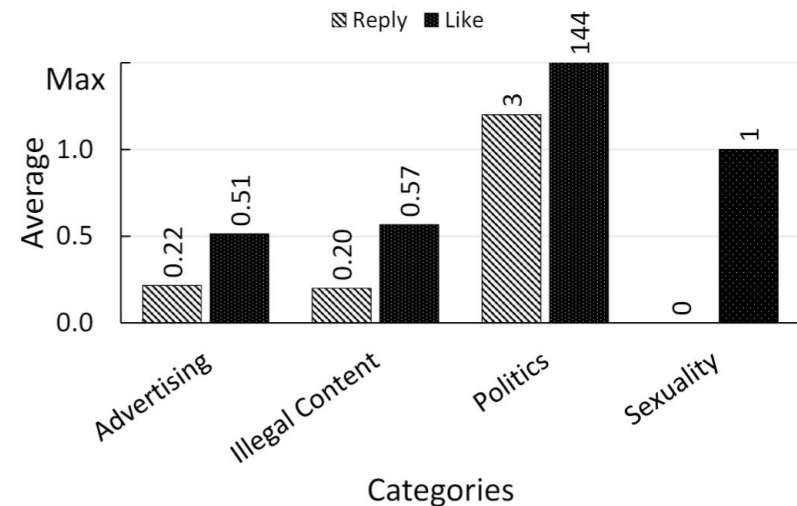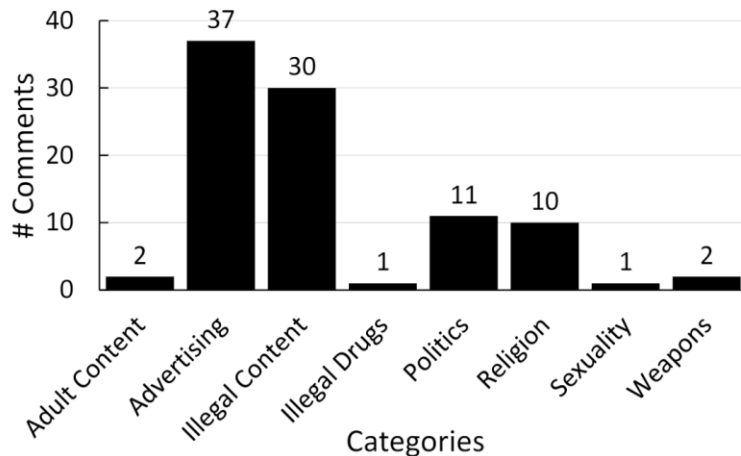# Kids Exposure to URLs

- We defined two metrics to estimate the prevalence and use of the URL by the audience.

  a. Video's popularity, represented by the number of views, likes, and comments.

 b. Comment's popularity, defined as the likes and replies on the comment containing the URL.
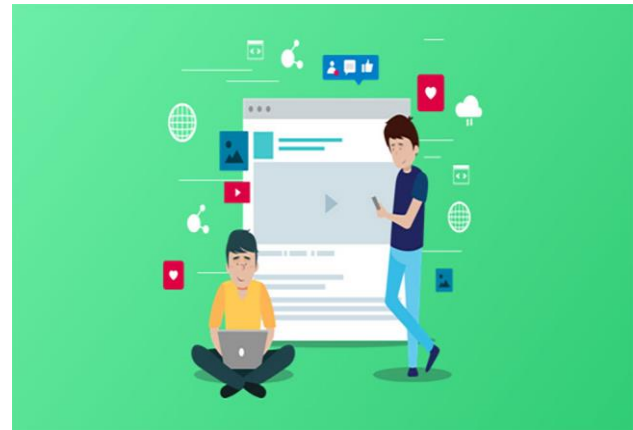
Kids Exposure to Inappropriate Topics

- Advertising and Illegal Content are popular within the URLs, with 71.27% of the total URLs.

- Comments with political URLs have on average three replies, and 144 likes.

Kids Exposure to Malicious URLs

- Kids from the age of 6 to 8 have the highest interaction with malicious URLs, represented as the average number of replies, likes, comments, and videos.
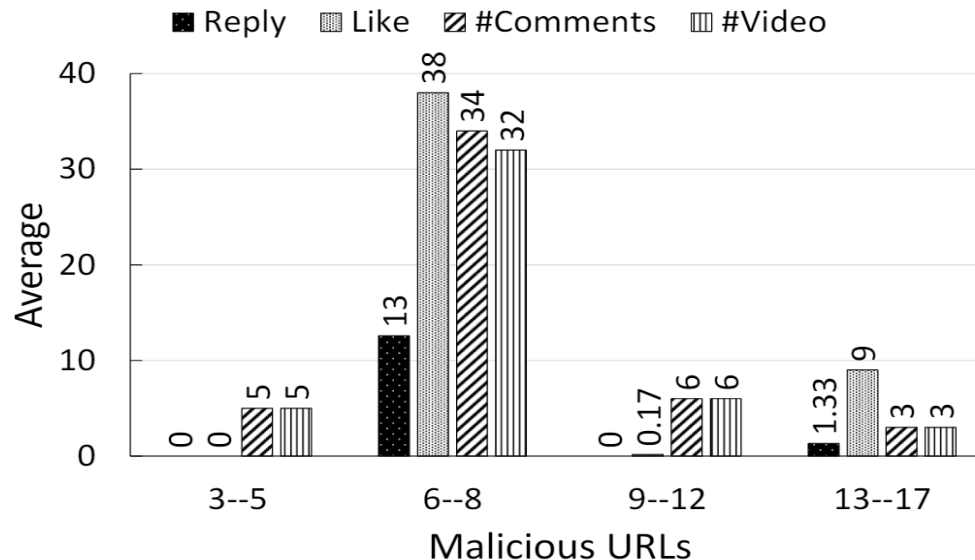
Kids Exposure to Malicious URLs

- Videos with malware sites URLs have an average number of viewers of more than 51 million views.

- More than 61million viewers of the videos with phishing.

- Higher number of viewers increases the likelihood of clicking on these links.

URLs Type	#Videos	Avg #viwers
Malicious site	47	46,061,532
Malware site	8	51,075,237
Phishing site	5	61,825,765

Conclusion Remarks

- We investigated the URLs embedded in comments on YouTube kids' videos, focusing on their content topic, and the presence of malicious URLs.

- Our findings highlight the exposure of kids to inappropriate and malicious URLs, calling for increased awareness of such exposure and take measures to ensure children's safety.

UCF

Thank you.

Contact information
Email: salshamrani@knights.ucf.edu